Uncertainty

Based on slides by S. Russell - http://aima.cs.berkeley.edu/

Uncertainty

Let's try to represent the knowledge that toothache is a symptom of cavities.

$Toothache \Rightarrow Cavity$

This states with absolute certainty that there is a cavity:
clearly inaccurate in the real world (teeth ache for many reasons).

Most things we know are uncertain:
some chance of failure - car might not start when it's -20 degrees
some degree of belief - "Leslie" is a girl (55 times out of 100)

Methods for handling uncertainty

Fuzzy logic: mostly handles vagueness

is it light or dark?

is it cold, moderate or warm? (each will have a degree of truth)

successful applications exist (camera auto exposure, train control)

Rules with fudge factors:

$Sprinkler \mapsto_{0.99} WetGrass$

$WetGrass \mapsto_{0.7} Rain$

Problem: Not clear how to compute certainties when combining rules

Probability

Given the available evidence,

the car starts with probability 0.95

Probability

Probability values **summarize** effects of

laziness: failure to enumerate exceptions, qualifications, etc.

ignorance: lack of relevant facts, initial conditions, etc.

Subjective or Bayesian probability:

e.g., $P(CarStarts|verycold) = 0.95$

Probabilities change with new evidence:

e.g., $P(CarStarts|verycold, oldbattery) = 0.25$

This is similar to

$VeryCold \wedge OldBattery \mapsto_{0.75} DoesNotStart$

but the probability approach allows using established, solid math

Probability basics

Begin with a set Ω—the sample space

e.g., 6 possible rolls of a die.

$\omega \in \Omega$ is a sample point/possible world/atomic event

A probability space or probability model is a sample space with an assignment $P(\omega)$ for every $\omega \in \Omega$ s.t.

$0 \leq P(\omega) \leq 1$

$\Sigma_\omega P(\omega) = 1$

e.g., $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$.

An event A is any subset of Ω

$$P(A) = \Sigma_{\{\omega \in A\}} P(\omega)$$

E.g., $P(\text{die roll} < 4) = P(1) + P(2) + P(3) = 1/6 + 1/6 + 1/6 = 1/2$

Random variables

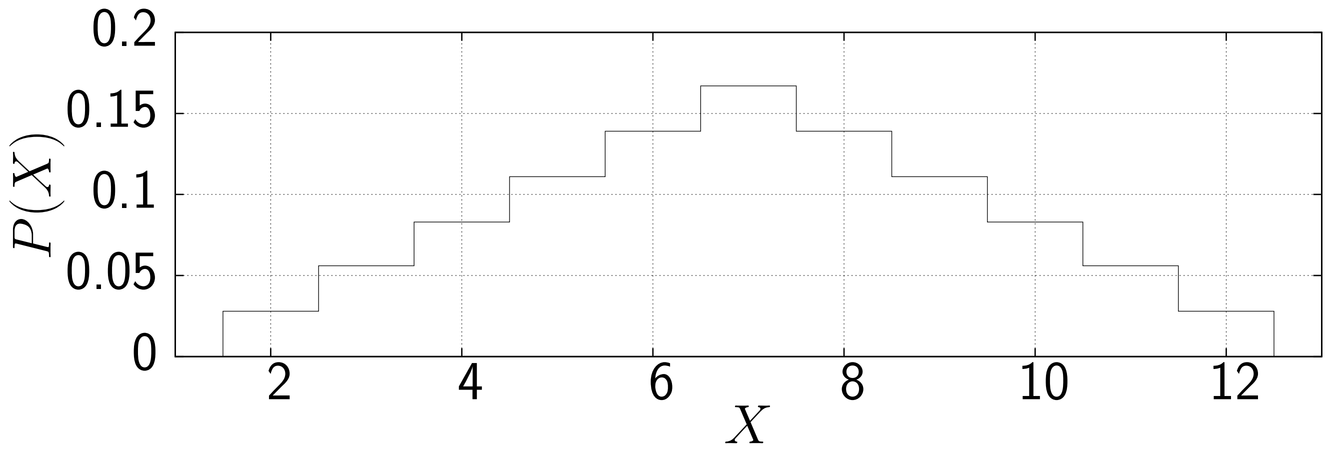
A random variable is a function from sample points to some range, e.g., the reals or Booleans

e.g., $Odd(1) = true$.

Each random variable has a probability distribution:

e.g., $P(Odd = true) = P(1) + P(3) + P(5) = 1/6 + 1/6 + 1/6 = 1/2$
$P(Odd = false) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 1/2$

Let X be a random discrete variable representing the total of two die rolls ($\Omega = \{(1,1),(1,2),\cdots,(6,5),(6,6)\}$). Probabilities of X values:



Conditional probability

Conditional or posterior probabilities

e.g., $P(Cavity = true | Toothache = true) = 0.8$

i.e., **given that** $Toothache = true$ **is all I know**

NOT "if toothache then 80% chance of cavity"

If we know more, e.g., $Cavity = true$ is also given, then we have

$$P(Cavity = true | Toothache = true, Cavity = true) = 1$$

New evidence may be irrelevant, allowing simplification, e.g.,

$$P(Cavity = true | Toothache = true, X = 7) = P(Cavity = true | Toothache = true)$$

Definition of conditional probability:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ if } P(b) \neq 0$$

Product rule gives an alternative formulation:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

Representing knowledge

Proposition is a statement which is either true or false.

Take the following **logical propositions**:

$Sprinkler$ (a fact)

$Rain \Rightarrow WetGrass$ (a rule)

$\neg Rain \wedge Sprinkler \Rightarrow WetGrass$

We state rules as conditional probabilities of a **proposition**, given **observations** (facts).

$P(WetGrass = true|Rain = true) = 0.98$

$P(WetGrass = true|Rain = false, Sprinkler = true) = 0.99$

Facts are used when computing the probability of a proposition. However, in case we do not have observations, background information can be included:

$P(Sprinkler = true) = 0.3$

Unlike logic, this is not a representation language. We are merely stating the probabilities of certain values for random variables.

Propositions

$Rain = false$ is itself a statement which is true or false
(compare to $Weather = sunny$), i.e. a **proposition**.

Think of a proposition as the event (set of sample points)
where the proposition is true

Given Boolean random variables A and B:

event a = set of sample points where $A(\omega) = true$
event $\neg a$ = set of sample points where $A(\omega) = false$
event $a \wedge b$ = points where $A(\omega) = true$ and $B(\omega) = true$

With Boolean variables, sample point = propositional logic model
e.g., $A = true$, $B = false$, or $a \wedge \neg b$.
Proposition = disjunction of atomic events in which it is true
e.g., $(a \vee b) \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b)$
$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + P(a \wedge b)$

Syntax for propositions

Propositional or Boolean random variables
  e.g., $Cavity$ (do I have a cavity?)
  $Cavity = true$ is a proposition, also written $cavity$

Discrete random variables (finite or infinite)
  e.g., $Weather$ is one of $\langle sunny, rain, cloudy, snow \rangle$
  $Weather = rain$ is a proposition
  Values must be exhaustive and mutually exclusive

Continuous random variables (bounded or unbounded)
  e.g., $Temp = 21.6$; also allow, e.g., $Temp < 22.0$.

Arbitrary Boolean combinations of basic propositions are also propositions.

Asking questions

How do we find out arbitrary probabilities? Given that the grass is wet and it is cloudy, what is the probability that we need an umbrella?

Start by recalling that $P(A)$ is the sum of probabilities of the sample points where A is true.

Often in AI applications, the sample points are **defined** by the values of a set of random variables, i.e., the sample space is the Cartesian product of the ranges of the variables

If our world (i. e. the way we model the domain we are working in) consists of discrete variables $Weather$ and $Cavity$ then we get 4×2 different combinations of their values. Each is a sample point ω.

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	ω_1	ω_2	ω_3	ω_4
$Cavity = false$	ω_5	ω_6	ω_7	ω_8

Prior probability

Prior or unconditional probabilities of propositions

e.g., $P(Cavity = true) = 0.2$ and $P(Weather = sunny) = 0.72$

correspond to belief prior to arrival of any (new) evidence

Probability distribution gives values for all possible assignments:

$\mathbf{P}(Weather) = \langle 0.72, 0.1, 0.08, 0.1 \rangle$ (normalized, i.e., sums to 1)

Joint probability distribution for a set of r.v.s gives the probability of every atomic event on those r.v.s (i.e., every sample point)

$\mathbf{P}(Weather, Cavity) =$ a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Every question about a domain can be answered by the joint distribution because every event is a sum of sample points

Inference by enumeration

Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$

Inference by enumeration

Start with the joint distribution:

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$

$P(toothache) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2$

Inference by enumeration

Start with the joint distribution:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

For any proposition ϕ, sum the atomic events where it is true:

$$P(\phi) = \Sigma_{\omega:\omega\models\phi} P(\omega)$$

$$P(cavity \lor toothache) = 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28$$

Inference by enumeration

Start with the joint distribution:

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	**.108**	**.012**	**.072**	**.008**
¬ cavity	**.016**	**.064**	**.144**	**.576**

Can also compute conditional probabilities:

$$P(\neg cavity | toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Inference by enumeration, contd.

The full joint distribution is similar to a truth table representing all possible models of a knowledge base in logic:

– gives theoretical justification for the approach
– impractical for real world applications

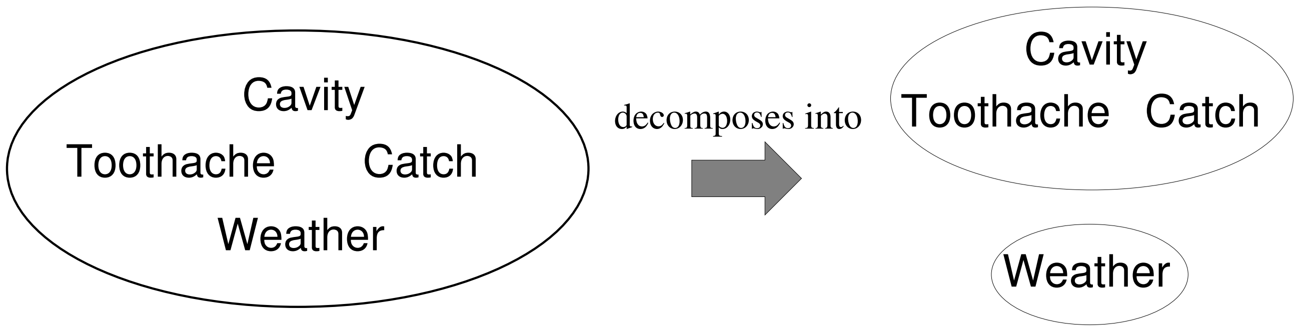
Obvious problems:

1) Worst-case time complexity $O(d^n)$ where d is the largest arity
2) Space complexity $O(d^n)$ to store the joint distribution
3) How to find the numbers for $O(d^n)$ entries???

Recall the $\mathbf{P}(Weather, Cavity)$ distribution. Any redundancy there?

Independence

A and B are independent iff
$\mathbf{P}(A|B) = \mathbf{P}(A)$ or $\mathbf{P}(B|A) = \mathbf{P}(B)$ or $\mathbf{P}(A, B) = \mathbf{P}(A)\mathbf{P}(B)$



$$\mathbf{P}(Toothache, Catch, Cavity, Weather)$$
$$= \mathbf{P}(Toothache, Catch, Cavity)\mathbf{P}(Weather)$$

32 entries reduced to 12.

Absolute independence powerful but rare

Dentistry is a large field with hundreds of variables,
none of which are independent. What to do?

Conditional distributions: background

Notation for conditional distributions: $\mathbf{P}(Cavity|Toothache)$ = 2-element vector of 2-element vectors)

$Toothache =$	$true$	$false$
$Cavity = true$	0.6	0.1
$Cavity = false$	0.4	0.9

Note the normalization by columns: $Toothache$ is already given here.

Product rule for distributions:

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather|Cavity)\mathbf{P}(Cavity)$$

(View as a 4×2 set of equations, **not** matrix mult.)

Chain rule is derived by successive application of product rule:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_1, \ldots, X_{n-1})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \mathbf{P}(X_1, \ldots, X_{n-2})\ \mathbf{P}(X_{n-1}|X_1, \ldots, X_{n-2})\ \mathbf{P}(X_n|X_1, \ldots, X_{n-1}) \\ &= \ldots \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1})\end{aligned}$$

Conditional independence

$\mathbf{P}(Toothache, Cavity, Catch)$ has $2^3 - 1 =$ 7 independent entries

If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:

(1) $P(catch|toothache, cavity) = P(catch|cavity)$

The same independence holds if I haven't got a cavity:

(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

$Catch$ is conditionally independent of $Toothache$ given $Cavity$:

$\mathbf{P}(Catch|Toothache, Cavity) = \mathbf{P}(Catch|Cavity)$

Equivalent statements:

$\mathbf{P}(Toothache|Catch, Cavity) = \mathbf{P}(Toothache|Cavity)$

$\mathbf{P}(Toothache, Catch|Cavity) = \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)$

Conditional independence contd.

Write out full joint distribution using chain rule:

$$\mathbf{P}(Toothache, Catch, Cavity)$$
$$= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch, Cavity)$$
$$= \mathbf{P}(Toothache|Catch, Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)$$
$$= \mathbf{P}(Toothache|Cavity)\mathbf{P}(Catch|Cavity)\mathbf{P}(Cavity)$$

I.e., $2 + 2 + 1 = 5$ independent numbers (equations 1 and 2 remove 2)

In most cases, the use of conditional independence reduces the size of the representation of the joint distribution from exponential in n to linear in n.

Conditional independence is our most basic and robust form of knowledge about uncertain environments.

Bayes' Rule

Product rule $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

$\Rightarrow$ Bayes' rule $P(a|b) = \dfrac{P(b|a)P(a)}{P(b)}$

or in distribution form (α is the normalization constant,
to make the sum of distribution 1).

$$\mathbf{P}(Y|X) = \frac{\mathbf{P}(X|Y)\mathbf{P}(Y)}{\mathbf{P}(X)} = \alpha \mathbf{P}(X|Y)\mathbf{P}(Y)$$

Useful for assessing diagnostic probability from causal probability:

$$P(Cause|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$$

E.g., let M be meningitis, S be stiff neck:

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)} = \frac{0.8 \times 0.0001}{0.1} = 0.0008$$

Bayes' Rule and conditional independence

$$
\begin{aligned}
&\mathbf{P}(Cavity|toothache \land catch) \\
&= \alpha\, \mathbf{P}(toothache \land catch|Cavity)\mathbf{P}(Cavity) \\
&= \alpha\, \mathbf{P}(toothache|Cavity)\mathbf{P}(catch|Cavity)\mathbf{P}(Cavity)
\end{aligned}
$$

This is an example of a naive Bayes model:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause)\Pi_i \mathbf{P}(Effect_i|Cause)$$

Total number of parameters is **linear** in n

Summary

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools